CLAIMS

What is claimed is:

1. A cost-adaptive cache, comprising:

a partitioned real cache, wherein data is stored in each the real cache partitions according to its replacement cost; and

a partitioned phantom cache to provide a directory of information pertaining to blocks of data which do not qualify for inclusion in the real cache,

whereby the partitions in the phantom cache correspond to the partitions in the real cache,

whereby the cost-adaptive cache maximizes performance in a system by preferentially caching data that is more costly to replace.

2. The cost-adaptive cache of claim 1 wherein the real cache comprises a variable number of blocks for storing data.

3. The cost-adaptive cache of claim 1 wherein the real cache includes a configurable number of partitions.

4. The cost-adaptive cache of claim 3 wherein the configurable number of partitions are identified according to a replacement cost of data included within each of the partitions.

5. The cost-adaptive cache of claim 3 wherein the partitions each have a pair of associated replacement cost values which define the boundaries for each of the partitions.

6. The cost-adaptive cache of claim 1 wherein the total size of corresponding partitions in the real cache and the phantom cache are less than or equal to the total size of the real cache.

7. The cost-adaptive cache of claim 1 wherein a target size is associated with each of the partitions in the real cache and the target sizes can be fixed, dynamic or adjusted periodically.

8. The cost-adaptive cache of claim 1 further comprises maintaining hit/miss statistics for each of the partitions in the real cache and each of the partitions in the phantom cache.

9. The cost-adaptive cache of claim 1 further comprises moving blocks between partitions within the real cache and the phantom cache in response to hits and misses in the caches.

10. The cost-adaptive cache of claim 9 further comprises adjusting the sizes of the partitions in the real cache to minimize the overall cost of servicing data requests, wherein the overall cost comprises the number of times data is requested and the cost of satisfying each of the requests.

11. The cost-adaptive cache of claim 10 wherein the adjustment is based on the hit/miss statistics of the partitions and the relative replacement cost of blocks in the partitions.

12. The cost-adaptive cache of claim 10 wherein the adjustment is based on the stack distance of a hit in the phantom cache.

13. The cost-adaptive cache of claim 9 wherein when a data block is evicted from the real cache, it is moved into the corresponding partition in the phantom cache.

14. The cost-adaptive cache of claim 13 wherein a block of data can be evicted from the phantom cache in order to make room for the data block evicted from the real cache.

15. The cost-adaptive cache of claim 1 wherein the replacement cost of a block of data is obtained by observing the length of time needed to service a request for that data.

16. A method for dynamically partitioning a storage system cache according to a replacement cost associated with data stored in the cache, the cache holding the data as blocks of data, the method comprising the steps of:

maintaining a history of recently evicted data blocks for each partition;

assigning data to one partition based on a cost associated with not keeping the data in the cache;

determining a future size for each partition based on the history and the cost associated with not keeping the data in the cache; and

whereby the cache's performance is dynamically maximized by preferentially caching data that are most costly to replace.

17. The method of claim 16 wherein the future size of each partition is determined so as to minimize the overall cost of servicing requests for data , wherein the overall cost comprises the number of times the data is requested and the cost of satisfying each of the requests.

18. The method of claim 16 wherein the cost of not keeping the data in the cache is obtained by observing the length of time needed to service a request for that data.